International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report **IR-05-079**

Evolution of Indirect Reciprocity

Martin A. Nowak (martin_nowak@harvard.edu)
Karl Sigmund (karl.sigmund@univie.ac.at)

Approved by

Ulf Dieckmann
Program Leader, ADN

December 2005

Interim Reports on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

IIASA STUDIES IN ADAPTIVE DYNAMICS NO. 111

The Adaptive Dynamics Network at IIASA fosters the development of new mathematical and conceptual techniques for understanding the evolution of complex adaptive systems.

Focusing on these long-term implications of adaptive processes in systems of limited growth, the Adaptive Dynamics Network brings together scientists and institutions from around the world with IIASA acting as the central node.

Scientific progress within the network is collected in the IIASA Studies in Adaptive Dynamics series.

No. 1 Metz JAJ, Geritz SAH, Meszéna G, Jacobs FJA, van Heerwaarden JS: *Adaptive Dynamics: A Geometrical Study of the Consequences of Nearly Faithful Reproduction.* IIASA Working Paper WP-95-099 (1995). van Strien SJ, Verduyn Lunel SM (eds): Stochastic and Spatial Structures of Dynamical Systems, Proceedings of the Royal Dutch Academy of Science (KNAW Verhandelingen), North Holland, Amsterdam, pp. 183-231 (1996).

No. 2 Dieckmann U, Law R: *The Dynamical Theory of Coevolution: A Derivation from Stochastic Ecological Processes.* IIASA Working Paper WP-96-001 (1996). Journal of Mathematical Biology 34:579-612 (1996).

No. 3 Dieckmann U, Marrow P, Law R: *Evolutionary Cycling of Predator-Prey Interactions: Population Dynamics and the Red Queen.* IIASA Preprint (1995). Journal of Theoretical Biology 176:91-102 (1995).

No. 4 Marrow P, Dieckmann U, Law R: *Evolutionary Dynamics of Predator-Prey Systems: An Ecological Perspective.* IIASA Working Paper WP-96-002 (1996). Journal of Mathematical Biology 34:556-578 (1996).

No. 5 Law R, Marrow P, Dieckmann U: *On Evolution under Asymmetric Competition.* IIASA Working Paper WP-96-003 (1996). Evolutionary Ecology 11:485-501 (1997).

No. 6 Metz JAJ, Mylius SD, Diekmann O: *When Does Evolution Optimize? On the Relation Between Types of Density Dependence and Evolutionarily Stable Life History Parameters.* IIASA Working Paper WP-96-004 (1996).

No. 7 Ferrière R, Gatto M: *Lyapunov Exponents and the Mathematics of Invasion in Oscillatory or Chaotic Populations.* Theoretical Population Biology 48:126-171 (1995).

No. 8 Ferrière R, Fox GA: *Chaos and Evolution.* IIASA Preprint (1996). Trends in Ecology and Evolution 10:480-485 (1995).

No. 9 Ferrière R, Michod RE: *The Evolution of Cooperation in Spatially Heterogeneous Populations.* IIASA Working Paper WP-96-029 (1996). The American Naturalist 147:692-717 (1996).

No. 10 van Dooren TJM, Metz JAJ: *Delayed Maturation in Temporally Structured Populations with Non-Equilibrium Dynamics.* IIASA Working Paper WP-96-070 (1996). Journal of Evolutionary Biology 11:41-62 (1998).

No. 11 Geritz SAH, Metz JAJ, Kisdi É, Meszéna G: *The Dynamics of Adaptation and Evolutionary Branching.* IIASA Working Paper WP-96-077 (1996). Physical Review Letters 78:2024-2027 (1997).

No. 12 Geritz SAH, Kisdi É, Meszéna G, Metz JAJ: *Evolutionary Singular Strategies and the Adaptive Growth and Branching of the Evolutionary Tree.* IIASA Working Paper WP-96-114 (1996). Evolutionary Ecology 12:35-57 (1998).

No. 13 Heino M, Metz JAJ, Kaitala V: *Evolution of Mixed Maturation Strategies in Semelparous Life-Histories: The Crucial Role of Dimensionality of Feedback Environment.* IIASA Working Paper WP-96-126 (1996). Philosophical Transactions of the Royal Society of London Series B 352:1647-1655 (1997).

No. 14 Dieckmann U: *Can Adaptive Dynamics Invade?* IIASA Working Paper WP-96-152 (1996). Trends in Ecology and Evolution 12:128-131 (1997).

No. 15 Meszéna G, Czibula I, Geritz SAH: *Adaptive Dynamics in a 2-Patch Environment: A Simple Model for Allopatric and Parapatric Speciation.* IIASA Interim Report IR-97-001 (1997). Journal of Biological Systems 5:265-284 (1997).

No. 16 Heino M, Metz JAJ, Kaitala V: *The Enigma of Frequency-Dependent Selection.* IIASA Interim Report IR-97-061 (1997). Trends in Ecology and Evolution 13:367-370 (1998).

No. 17 Heino M: *Management of Evolving Fish Stocks.* IIASA Interim Report IR-97-062 (1997). Canadian Journal of Fisheries and Aquatic Sciences 55:1971-1982 (1998).

No. 18 Heino M: *Evolution of Mixed Reproductive Strategies in Simple Life-History Models.* IIASA Interim Report IR-97-063 (1997).

No. 19 Geritz SAH, van der Meijden E, Metz JAJ: *Evolutionary Dynamics of Seed Size and Seedling Competitive Ability.* IIASA Interim Report IR-97-071 (1997). Theoretical Population Biology 55:324-343 (1999).

No. 20 Galis F, Metz JAJ: *Why Are There So Many Cichlid Species? On the Interplay of Speciation and Adaptive Radiation.* IIASA Interim Report IR-97-072 (1997). Trends in Ecology and Evolution 13:1-2 (1998).

No. 21 Boerlijst MC, Nowak MA, Sigmund K: *Equal Pay for all Prisoners/ The Logic of Contrition.* IIASA Interim Report IR-97-073 (1997). American Mathematical Society Monthly 104:303-307 (1997). Journal of Theoretical Biology 185:281-293 (1997).

No. 22 Law R, Dieckmann U: *Symbiosis Without Mutualism and the Merger of Lineages in Evolution.* IIASA Interim Report IR-97-074 (1997). Proceedings of the Royal Society of London Series B 265:1245-1253 (1998).

No. 23 Klinkhamer PGL, de Jong TJ, Metz JAJ: *Sex and Size in Cosexual Plants.* IIASA Interim Report IR-97-078 (1997). Trends in Ecology and Evolution 12:260-265 (1997).

No. 24 Fontana W, Schuster P: *Shaping Space: The Possible and the Attainable in RNA Genotype-Phenotype Mapping.* IIASA Interim Report IR-98-004 (1998). Journal of Theoretical Biology 194:491-515 (1998).

No. 25 Kisdi É, Geritz SAH: *Adaptive Dynamics in Allele Space: Evolution of Genetic Polymorphism by Small Mutations in a Heterogeneous Environment.* IIASA Interim Report IR-98-038 (1998). Evolution 53:993-1008 (1999).

No. 26 Fontana W, Schuster P: *Continuity in Evolution: On the Nature of Transitions.* IIASA Interim Report IR-98-039 (1998). Science 280:1451-1455 (1998).

No. 27 Nowak MA, Sigmund K: *Evolution of Indirect Reciprocity by Image Scoring/ The Dynamics of Indirect Reciprocity.* IIASA Interim Report IR-98-040 (1998). Nature 393:573-577 (1998). Journal of Theoretical Biology 194:561-574 (1998).

No. 28 Kisdi É: *Evolutionary Branching Under Asymmetric Competition.* IIASA Interim Report IR-98-045 (1998). Journal of Theoretical Biology 197:149-162 (1999).

No. 29 Berger U: *Best Response Adaptation for Role Games.* IIASA Interim Report IR-98-086 (1998).

No. 30 van Dooren TJM: *The Evolutionary Ecology of Dominance-Recessivity.* IIASA Interim Report IR-98-096 (1998). Journal of Theoretical Biology 198:519-532 (1999).

No. 31 Dieckmann U, O'Hara B, Weisser W: *The Evolutionary Ecology of Dispersal.* IIASA Interim Report IR-98-108 (1998). Trends in Ecology and Evolution 14:88-90 (1999).

No. 32 Sigmund K: *Complex Adaptive Systems and the Evolution of Reciprocation.* IIASA Interim Report IR-98-100 (1998). Ecosystems 1:444-448 (1998).

No. 33 Posch M, Pichler A, Sigmund K: *The Efficiency of Adapting Aspiration Levels.* IIASA Interim Report IR-98-103 (1998). Proceedings of the Royal Society London Series B 266:1427-1435 (1999).

No. 34 Mathias A, Kisdi É: *Evolutionary Branching and Coexistence of Germination Strategies.* IIASA Interim Report IR-99-014 (1999).

No. 35 Dieckmann U, Doebeli M: *On the Origin of Species by Sympatric Speciation.* IIASA Interim Report IR-99-013 (1999). Nature 400:354-357 (1999).

No. 36 Metz JAJ, Gyllenberg M: *How Should We Define Fitness in Structured Metapopulation Models? Including an Application to the Calculation of Evolutionarily Stable Dispersal Strategies.* IIASA Interim Report IR-99-019 (1999). Proceedings of the Royal Society of London Series B 268:499-508 (2001).

No. 37 Gyllenberg M, Metz JAJ: *On Fitness in Structured Metapopulations.* IIASA Interim Report IR-99-037 (1999). Journal of Mathematical Biology 43:545-560 (2001).

No. 38 Meszéna G, Metz JAJ: *Species Diversity and Population Regulation: The Importance of Environmental Feedback Dimensionality.* IIASA Interim Report IR-99-045 (1999).

No. 39 Kisdi É, Geritz SAH: *Evolutionary Branching and Sympatric Speciation in Diploid Populations.* IIASA Interim Report IR-99-048 (1999).

No. 40 Ylikarjula J, Heino M, Dieckmann U: *Ecology and Adaptation of Stunted Growth in Fish.* IIASA Interim Report IR-99-050 (1999). Evolutionary Ecology 13:433-453 (1999).

No. 41 Nowak MA, Sigmund K: *Games on Grids.* IIASA Interim Report IR-99-038 (1999). Dieckmann U, Law R, Metz JAJ (eds): The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, Cambridge, UK, pp. 135-150 (2000).

No. 42 Ferrière R, Michod RE: *Wave Patterns in Spatial Games and the Evolution of Cooperation.* IIASA Interim Report IR-99-041 (1999). Dieckmann U, Law R, Metz JAJ (eds): The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, Cambridge, UK, pp. 318-332 (2000).

No. 43 Kisdi É, Jacobs FJA, Geritz SAH: *Red Queen Evolution by Cycles of Evolutionary Branching and Extinction.* IIASA Interim Report IR-00-030 (2000). Selection 2:161-176 (2001).

No. 44 Meszéna G, Kisdi É, Dieckmann U, Geritz SAH, Metz JAJ: *Evolutionary Optimisation Models and Matrix Games in the Unified Perspective of Adaptive Dynamics.* IIASA Interim Report IR-00-039 (2000). Selection 2:193-210 (2001).

No. 45 Parvinen K, Dieckmann U, Gyllenberg M, Metz JAJ: *Evolution of Dispersal in Metapopulations with Local Density Dependence and Demographic Stochasticity.* IIASA Interim Report IR-00-035 (2000). Journal of Evolutionary Biology 16:143-153 (2003).

No. 46 Doebeli M, Dieckmann U: *Evolutionary Branching and Sympatric Speciation Caused by Different Types of Ecological Interactions.* IIASA Interim Report IR-00-040 (2000). The American Naturalist 156:S77-S101 (2000).

No. 47 Heino M, Hanski I: *Evolution of Migration Rate in a Spatially Realistic Metapopulation Model.* IIASA Interim Report IR-00-044 (2000). The American Naturalist 157:495-511 (2001).

No. 48 Gyllenberg M, Parvinen K, Dieckmann U: *Evolutionary Suicide and Evolution of Dispersal in Structured Metapopulations.* IIASA Interim Report IR-00-056 (2000). Journal of Mathematical Biology 45:79-105 (2002).

No. 49 van Dooren TJM: *The Evolutionary Dynamics of Direct Phenotypic Overdominance: Emergence Possible, Loss Probable.* IIASA Interim Report IR-00-048 (2000). Evolution 54:1899-1914 (2000).

No. 50 Nowak MA, Page KM, Sigmund K: *Fairness Versus Reason in the Ultimatum Game.* IIASA Interim Report IR-00-57 (2000). Science 289:1773-1775 (2000).

No. 51 de Feo O, Ferrière R: *Bifurcation Analysis of Population Invasion: On-Off Intermittency and Basin Riddling.* IIASA Interim Report IR-00-074 (2000). International Journal of Bifurcation and Chaos 10:443-452 (2000).

No. 52 Heino M, Laaka-Lindberg S: *Clonal Dynamics and Evolution of Dormancy in the Leafy Hepatic Lophozia Silvicola.* IIASA Interim Report IR-01-018 (2001). Oikos 94:525-532 (2001).

No. 53 Sigmund K, Hauert C, Nowak MA: *Reward and Punishment in Minigames.* IIASA Interim Report IR-01-031 (2001). Proceedings of the National Academy of Sciences of the USA 98:10757-10762 (2001).

No. 54 Hauert C, De Monte S, Sigmund K, Hofbauer J: *Oscillations in Optional Public Good Games.* IIASA Interim Report IR-01-036 (2001).

No. 55 Ferrière R, Le Galliard J: *Invasion Fitness and Adaptive Dynamics in Spatial Population Models.* IIASA Interim Report IR-01-043 (2001). Clobert J, Dhondt A, Danchin E, Nichols J (eds): Dispersal, Oxford University Press, pp. 57-79 (2001).

No. 56 de Mazancourt C, Loreau M, Dieckmann U: *Can the Evolution of Plant Defense Lead to Plant-Herbivore Mutualism.* IIASA Interim Report IR-01-053 (2001). The American Naturalist 158:109-123 (2001).

No. 57 Claessen D, Dieckmann U: *Ontogenetic Niche Shifts and Evolutionary Branching in Size-Structured Populations.* IIASA Interim Report IR-01-056 (2001). Evolutionary Ecology Research 4:189-217 (2002).

No. 58 Brandt H: *Correlation Analysis of Fitness Landscapes.* IIASA Interim Report IR-01-058 (2001).

No. 59 Dieckmann U: *Adaptive Dynamics of Pathogen-Host Interacations.* IIASA Interim Report IR-02-007 (2002). Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds): Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management, Cambridge University Press, Cambridge, UK, pp. 39-59 (2002).

No. 60 Nowak MA, Sigmund K: *Super- and Coinfection: The Two Extremes.* IIASA Interim Report IR-02-008 (2002). Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds): Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management, Cambridge University Press, Cambridge, UK, pp. 124-137 (2002).

No. 61 Sabelis MW, Metz JAJ: *Taking Stock: Relating Theory to Experiment.* IIASA Interim Report IR-02-009 (2002). Dieckmann U, Metz JAJ, Sabelis MW, Sigmund K (eds): Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management, Cambridge University Press, Cambridge, UK, pp. 379-398 (2002).

No. 62 Cheptou P, Dieckmann U: *The Evolution of Self-Fertilization in Density-Regulated Populations* . IIASA Interim Report IR-02-024 (2002). Proceedings of the Royal Society of London Series B 269:1177-1186 (2002).

No. 63 Bürger R: *Additive Genetic Variation Under Intraspecific Competition and Stabilizing Selection: A Two-Locus Study.* IIASA Interim Report IR-02-013 (2002). Theoretical Population Biology 61:197-213 (2002).

No. 64 Hauert C, De Monte S, Hofbauer J, Sigmund K: *Volunteering as Red Queen Mechanism for Co-operation in Public Goods Games.* IIASA Interim Report IR-02-041 (2002). Science 296:1129-1132 (2002).

No. 65 Dercole F, Ferrière R, Rinaldi S: *Ecological Bistability and Evolutionary Reversals under Asymmetrical Competition.* IIASA Interim Report IR-02-053 (2002). Evolution 56:1081-1090 (2002).

No. 66 Dercole F, Rinaldi S: *Evolution of Cannibalistic Traits: Scenarios Derived from Adaptive Dynamics.* IIASA Interim Report IR-02-054 (2002). Theoretical Population Biology 62:365-374 (2002).

No. 67 Bürger R, Gimelfarb A: *Fluctuating Environments and the Role of Mutation in Maintaining Quantitative Genetic Variation.* IIASA Interim Report IR-02-058 (2002). Genetical Research 80:31-46 (2002).

No. 68 Bürger R: *On a Genetic Model of Intraspecific Competition and Stabilizing Selection.* IIASA Interim Report IR-02-062 (2002). Amer. Natur. 160:661-682 (2002).

No. 69 Doebeli M, Dieckmann U: *Speciation Along Environmental Gradients.* IIASA Interim Report IR-02-079 (2002). Nature 421:259-264 (2003).

No. 70 Dercole F, Irisson J, Rinaldi S: *Bifurcation Analysis of a Prey-Predator Coevolution Model.* IIASA Interim Report IR-02-078 (2002). SIAM Journal on Applied Mathematics 63:1378-1391 (2003).

No. 71 Le Galliard J, Ferrière R, Dieckmann U: *The Adaptive Dynamics of Altruism in Spatially Heterogeneous Populations.* IIASA Interim Report IR-03-006 (2003). Evolution 57:1-17 (2003).

No. 72 Taborsky B, Dieckmann U, Heino M: *Unexpected Discontinuities in Life-History Evolution under Size-Dependent Mortality.* IIASA Interim Report IR-03-004 (2003). Proceedings of the Royal Society of London Series B 270:713-721 (2003).

No. 73 Gardmark A, Dieckmann U, Lundberg P: *Life-History Evolution in Harvested Populations: The Role of Natural Predation.* IIASA Interim Report IR-03-008 (2003). Evolutionary Ecology Research 5:239-257 (2003).

No. 74 Mizera F, Meszéna G: *Spatial Niche Packing, Character Displacement and Adaptive Speciation Along an Environmental Gradient.* IIASA Interim Report IR-03-062 (2003). Evolutionary Ecology Research 5:363-382 (2003).

No. 75 Dercole F: *Remarks on Branching-Extinction Evolutionary Cycles.* IIASA Interim Report IR-03-077 (2003). Journal of Mathematical Biology 47:569-580 (2003).

No. 76 Hofbauer J, Sigmund K: *Evolutionary Game Dynamics.* IIASA Interim Report IR-03-078 (2003). Bulletin of the American Mathematical Society 40:479-519 (2003).

No. 77 Ernande B, Dieckmann U, Heino M: *Adaptive Changes in Harvested Populations: Plasticity and Evolution of Age and Size at Maturation.* IIASA Interim Report IR-03-058 (2003). Proceedings of the Royal Society of London Series B-Biological Sciences 271:415-423 (2004).

No. 78 Hanski I, Heino M: *Metapopulation-Level Adaptation of Insect Host Plant Preference and Extinction-Colonization Dynamics in Heterogeneous Landscapes.* IIASA Interim Report IR-03-028 (2003). Theoretical Population Biology 63:309-338 (2003).

No. 79 van Doorn G, Dieckmann U, Weissing FJ: *Sympatric Speciation by Sexual Selection: A Critical Re-Evaluation.* IIASA Interim Report IR-04-003 (2004). American Naturalist 163:709-725 (2004).

No. 80 Egas M, Dieckmann U, Sabelis MW: *Evolution Restricts the Coexistence of Specialists and Generalists - the Role of Trade-off Structure.* IIASA Interim Report IR-04-004 (2004). American Naturalist 163:518-531 (2004).

No. 81 Ernande B, Dieckmann U: *The Evolution of Phenotypic Plasticity in Spatially Structured Environments: Implications of Intraspecific Competition, Plasticity Costs, and Environmental Characteristics.* IIASA Interim Report IR-04-006 (2004). Journal of Evolutionary Biology 17:613-628 (2004).

No. 82 Cressman R, Hofbauer J: *Measure Dynamics on a One-Dimensional Continuous Trait Space: Theoretical Foundations for Adaptive Dynamics.* IIASA Interim Report IR-04-016 (2004).

No. 83 Cressman R: *Dynamic Stability of the Replicator Equation with Continuous Strategy Space.* IIASA Interim Report IR-04-017 (2004).

No. 84 Ravigné V, Olivieri I, Dieckmann U: *Implications of Habitat Choice for Protected Polymorphisms.* IIASA Interim Report IR-04-005 (2004). Evolutionary Ecology Research 6:125-145 (2004).

No. 85 Nowak MA, Sigmund K: *Evolutionary Dynamics of Biological Games.* IIASA Interim Report IR-04-013 (2004). Science 303:793-799 (2004).

No. 86 Vukics A, Asbóth J, Meszéna G: *Speciation in Multidimensional Evolutionary Space.* IIASA Interim Report IR-04-028 (2004). Physical Review 68:041-903 (2003).

No. 87 de Mazancourt C, Dieckmann U: *Trade-off Geometries and Frequency-dependent Selection.* IIASA Interim Report IR-04-039 (2004). American Naturalist 164:765-778 (2004).

No. 88 Cadet CR, Metz JAJ, Klinkhamer PGL: *Size and the Not-So-Single Sex: disentangling the effects of size on sex allocation.* IIASA Interim Report IR-04-084 (2004). American Naturalist 164:779-792 (2004).

No. 89 Rueffler C, van Dooren TJM, Metz JAJ: *Adaptive Walks on Changing Landscapes: Levins' Approach Extended.* IIASA Interim Report IR-04-083 (2004). Theoretical Population Biology 65:165-178 (2004).

No. 90 de Mazancourt C, Loreau M, Dieckmann U: *Understanding Mutualism When There is Adaptation to the Partner.* IIASA Interim Report IR-05-016 (2005). Journal of Ecology 93:305-314 (2005).

No. 91 Dieckmann U, Doebeli M: *Pluralism in Evolutionary Theory.* IIASA Interim Report IR-05-017 (2005). Journal of Evolutionary Biology 18:1209-1213 (2005).

No. 92 Doebeli M, Dieckmann U, Metz JAJ, Tautz D: *What We Have Also Learned.* IIASA Interim Report IR-05-018 (2005). Evolution 59:691-695 (2005).

No. 93 Egas M, Sabelis MW, Dieckmann U: *Evolution of Specialization and Ecological Character Displacement of Herbivores Along a Gradient of Plant Quality.* IIASA Interim Report IR-05-019 (2005). Evolution 59:507-520 (2005).

No. 94 Le Galliard J, Ferrière R, Dieckmann U: *Adaptive Evolution of Social Traits: Origin, Trajectories, and Correlations of Altruism and Mobility.* IIASA Interim Report IR-05-020 (2005). American Naturalist 165:206-224 (2005).

No. 95 Doebeli M, Dieckmann U: *Adaptive Dynamics as a Mathematical Tool for Studying the Ecology of Speciation Processes.* IIASA Interim Report IR-05-022 (2005). Journal of Evolutionary Biology 18:1194-1200 (2005).

No. 96 Brandt H, Sigmund K: *The Logic of Reprobation: Assessment and Action Rules for Indirect Reciprocity.* IIASA Interim Report IR-04-085 (2004). Journal of Theoretical Biology 231:475-486 (2004).

No. 97 Hauert C, Haiden N, Sigmund K: *The Dynamics of Public Goods.* IIASA Interim Report IR-04-086 (2004). Discrete and Continuous Dynamical Systems - Series B 4:575-587 (2004).

No. 98 Meszéna G, Gyllenberg M, Jacobs FJA, Metz JAJ: *Link Between Population Dynamics and Dynamics of Darwinian Evolution.* IIASA Interim Report IR-05-026 (2005). Physical Review Letters 95:Article 078105 (2005).

No. 99 Meszéna G: *Adaptive Dynamics: The Continuity Argument.* IIASA Interim Report IR-05-032 (2005). Journal of Evolutionary Biology 18:1182-1185 (2005).

No. 100 Brännström NA, Dieckmann U: *Evolutionary Dynamics of Altruism and Cheating Among Social Amoebas.* IIASA Interim Report IR-05-039 (2005). Proceedings of the Royal Society London Series B 272:1609-1616 (2005).

No. 101 Meszéna G, Gyllenberg M, Pasztor L, Metz JAJ: *Competitive Exclusion and Limiting Similarity: A Unified Theory.* IIASA Interim Report IR-05-040 (2005).

No. 102 Szabo P, Meszéna G: *Limiting Similarity Revisited.* IIASA Interim Report IR-05-050 (2005).

No. 103 Krakauer DC, Sasaki A: *The Greater than Two-Fold Cost of Integration for Retroviruses.* IIASA Interim Report IR-05-069 (2005).

No. 104 Metz JAJ: *Eight Personal Rules for Doing Science.* IIASA Interim Report IR-05-069 (2005). Journal of Evolutionary Biology 18:1178-1181 (2005).

No. 105 Beltman JB, Metz JAJ: *Speciation: More Likely Through a Genetic or Through a Learned Habitat Preference?* IIASA Interim Report IR-05-072 (2005). Proceedings of the Royal Society of London Series B 272:1455-1463 (2005).

No. 106 Durinx M, Metz JAJ: *Multi-type Branching Processes and Adaptive Dynamics of Structured Populations.* IIASA Interim Report IR-05-065 (2005). Haccou P, Jager P, Vatutin V (eds): Branching Processes: Variation, Growth and Extinction of Populations, Cambridge University Press, Cambridge, UK, pp. 266-278 (2005).

No. 107 Brandt H, Sigmund K: *The Good, the Bad and the Discriminator - Errors in Direct and Indirect Reciprocity.* IIASA Interim Report IR-05-070 (2005).

No. 108 Brandt H, Sigmund K: *Punishing and Abstaining for Public Goods.* IIASA Interim Report IR-05-071 (2005).

No. 109 Ohtsuki A, Sasaki A: *Epidemiology and Disease-Control Under Gene-for-Gene Plant-Pathogen Interaction.* IIASA Interim Report IR-05-068 (2005).

No. 110 Brandt H, Sigmund K: *Indirect Reciprocity, Image-Scoring, and Moral Hazard.* IIASA Interim Report IR-05-078 (2005). Proceedings of the National Academy of Sciences of the United States of America 102:2666-2670 (2005).

No. 111 Nowak MA, Sigmund K: *Evolution of Indirect Reciprocity.* IIASA Interim Report IR-05-079 (2005). Nature 437:1292-1298 (2005).

Issues of the IIASA Studies in Adaptive Dynamics series can be obtained at www.iiasa.ac.at/Research/ADN/Series.html or by writing to adn@iiasa.ac.at.

Contents

Abstract 1
Introduction 1
Direct versus indirect reciprocity 2
Binary assessment, or the world in black and white 4
Costly experiments 5
Public goods 7
Cerebral expansion 8
Box 1: The good, the bad and the discriminating 8
Box 2: Let a thousand morals bloom 9
Box 3: Bidding for trust 10
Box 4: Games of direct reciprocity 10
Box 5: Upstream and downstream 11
Box 6: Social viscosity 11
References 12
Figure legends 16
Figures 17

Evolution of indirect reciprocity

Martin A. Nowak[1] & Karl Sigmund[2]

[1]Program for Evolutionary Dynamics, Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge, MA 02138, USA

[2]Faculty for Mathematics, University of Vienna, A-1090 Vienna, Austria
and IIASA, Laxenburg, Austria

Natural selection favours the strong and selfish who maximize their own resources at the expense of others. But many biological systems, and especially human societies, are organized around altruistic, cooperative interactions. How can natural selection promote unselfish behaviour? Various mechanisms have been proposed including kin selection, group selection and reciprocal altruism. The latter is traditionally formulated as direct reciprocity: I help you and you help me. More recently, a rich analysis of indirect reciprocity has emerged: I help you and somebody else helps me. Direct reciprocity relies on one's own experience in past encounters, whereas indirect reciprocity also uses the experience of others. The evolution of cooperation by indirect reciprocity leads to reputation building, morality judgement and complex social interactions with ever increasing cognitive demands.

Humans are the champions of reciprocity. Experiments and everyday experience alike show that what Adam Smith called 'our instinct to trade, barter and truck' relies to a considerable part on the widespread tendency to return helpful or harmful acts in kind. We do so even if these acts have been directed not to us but to others. This has been analysed under the headings of 'third party altruism'[1] or 'indirect reciprocity'[2], and has led to a considerable amount of experimental and theoretical investigations over the last few years.

Direct reciprocity is captured in the principle: 'You scratch my back, and I'll scratch yours'. But it is harder to make sense of the principle 'You scratch my back and I'll scratch someone else's.'[3]. Why should I do so? Presumably because I will not get scratched if it becomes known that I scratch nobody. Indirect reciprocity, in this view, is based on reputation. But why should anyone care about what I did to a third?

There are two approaches converging on this issue. One is rooted in social science, the other in evolutionary biology.

The main reason why economists and social scientists are interested in indirect reciprocity is because one-shot interactions between anonymous partners in a global market become increasingly frequent and tend to replace the traditional long-lasting associations and long-term interactions between relatives, neighbours, or members of the same village. A substantial part of our life is spent in the company of strangers[4], and many transactions are no longer face-to-face. The growth of e-auctions and other forms of e-commerce is based, to a considerable degree, on reputation and trust[5-10]. The possibility to exploit such trust raises what economists call moral hazards. How effective is reputation, especially if information is only partial?

Evolutionary biologists, on the other hand, are interested in the emergence of human societies, which constitutes the last (up to now) of the major transitions in evolution[11]. In contrast to other eusocial species, such as bees, ants or termites, humans display a large amount of

cooperation between non-relatives[12-14]. A considerable part of human cooperation is based on moralistic emotions – for instance, anger directed towards cheaters, or the proverbial 'warm inner glow' felt after performing an altruistic action. Neuro-economic experiments show that these emotions have a physiological basis[15-17]. Intriguingly, humans not only feel strongly about interactions which involve them directly, they also judge the actions between third parties, as evidenced by the contents of gossip[18]. Thus indirect reciprocity is likely to be connected with the origins of moral norms. Such norms are evidently to a large extent culture-specific; but the capacity for moral norms appears to be a human universal for which there is little evidence in other species[19].

Because the recent rapid advance of experimental investigations of indirect reciprocity was to a large part theory-driven, we shall discuss the modelling approaches before reviewing the experiments. But first we note, in a wider context, that indirect reciprocity seems to require a 'theory of mind'[20]. Whereas altruism directed towards kin works because similar genomes exist in different organisms, reciprocal altruism recognises that similar minds reside in different brains. It is easy to conceive that an organism experiences as 'good' or 'bad' anything that affects the own reproductive fitness in a positive or negative sense. The step from there to judging, as 'good' or 'bad', actions between third parties, is not obvious. The same terms 'good' and 'bad' which are applied to pleasure and pain are also used for moral judgements: this linguistic quirk reveals an astonishing degree of empathy, and reflects highly developed faculties for cognition and abstraction.

Direct versus indirect reciprocity

In the terminology based on Hamilton, Trivers, and Wilson[12-14], an act is said to be altruistic if it is costly to perform, but confers a benefit to another individual. In evolutionary biology, costs and benefits are measured in Darwinian fitness, which means reproductive success. In other contexts other utility scales such as monetary rewards may be more appropriate. Reciprocal altruism in its original, 'direct' sense is defined as an exchange of altruistic acts between the same two individuals, so that in total, both obtain a net benefit[1]. In the simplest model, the altruistic act consists in conferring a benefit b to the recipient at a cost c to the donor. We shall always assume that the cost is smaller than the benefit, so that if the act is returned, both individuals experience a gain. This yields the payoff structure of the familiar Prisoner's Dilemma game[21]. If both players cooperate, each receives b-c, which is better than what they would obtain by defecting, namely 0. But a unilateral defector would earn b, which is the highest payoff, and the exploited co-operator would pay the cost c without receiving any benefit. The payoff-maximising move consists in defecting (Fig 1).

This changes if the game is repeated for several rounds. For simplicity, we shall assume that in each round both players decide simultaneously. We could also assume that they alternate, which leads to a slightly different game[22-24]. The so-called folk theorem on repeated games states that if the probability for future rounds is sufficiently high, cooperation can be sustained by so-called trigger strategies, which switch to relentless defection as soon as the co-player defects once[25]. A rational player would have to weigh the benefit of exploiting the co-player in one round against the cost of forfeiting collaboration in all future rounds, and thus would abstain from defection.

In the context of indirect reciprocity, any two players are supposed to interact at most once with each other. Thus it is not possible that a cheater is taken to account by the victim. Each player can experience many rounds, but never with the same partner twice. Clearly, trigger

strategies can still ensure a cooperative Nash equilibrium, such that if all players use them, no player would have an incentive to deviate. In strategic thinking, only the payoffs matter, not by whom they are provided. In this sense, the step from direct to indirect reciprocity corresponds simply to the step from personal enforcement to community enforcement[26-29]. But a trigger strategy prescribing each person to cooperate until the first defection is personally experienced, and thenceforth to defect, hurts the original wrong-doer only after many rounds. A strategy triggered by the first defection in the population leaves cooperation at the mercy of the first wrong move. In both cases a lot of innocents would be punished, and errors cause havoc. Obviously, retaliation should be directed towards the cheater, rather than towards the whole community. This requires more detailed information. Game theory shows that even if information is transmitted only locally and errors occur occasionally, cooperation can be sustained: there exist strategies such that no rational player has an interest to deviate unilaterally[27].

In evolutionary game theory, it is not assumed that players are rational, but only that successful strategies spread -- by being inherited, for instance, or copied through imitation or learning[30]. For direct reciprocity, game theoretical analysis and individual-based simulations have shown that a population of defectors can be invaded by a small cluster of retaliators[31] or even by a single retaliator[32]. Typically, one considers a well-mixed population where individuals meet randomly and play a series of Prisoner's Dilemma games with each other. What counts is the total payoff. Retaliators compensate the loss of being exploited by a defector in the first round with long sequences of altruistic exchange with other retaliators. Once cooperation is established, a complex evolution takes place, which depends on the size of the population, the cost-to-benefit ratio, the average number of rounds and the probability of errors[31,33,34].

A similar model of indirect reciprocity assumes that within a well-mixed population, individuals meet randomly, one in the role of the potential donor, the other as potential recipient (Fig 2). Each individual experiences several rounds of this interaction in both roles, but never with the same partner twice. Again, all that counts is the total payoff. A player can follow either an unconditional strategy, such as always cooperate or always defect, or else a conditional strategy, which discriminates among the potential recipients according to past interactions. In a simple example, a discriminating player can help the co-player if that co-player's score exceeds a certain threshold. A player's score is 0 at birth, increases whenever that player helps and decreases whenever the player withholds help. Individual-based simulations show that if the cost-to-benefit ratio is sufficiently low, and the amount of information about the co-player's past sufficiently high, cooperation based on discrimination can emerge. In the resulting population, help is channelled towards those who help[35-37].

Two features of this model were immediately apparent[35,38,39]. One is the paradoxical nature of the discriminating strategy. In terms of rational game theory, why should players care about the scores of others rather than just about their own payoff, and why should they decrease their own score (and thus their likelihood of receiving help on later occasions) by withholding help from low-scorers? The second issue concerns the lack of stability of the cooperative outcome. The simulations display occasional bursts of defection, which are based on a previous build-up of undiscriminating altruists. In a population of discriminators, unconditional cooperators can increase by random drift and eventually invite the invasion of defectors.

The two issues are closely related. Intuitively speaking, the intentions of players who selfishly care about their own income only, and who give help, accordingly, just to keep their own

score high, are vastly different from the intentions of altruists who have only the interests of their co-players in mind and help them on every occasion. But the effect is the same in both cases: support will be given regardless of the co-players' contributions.

Binary assessment, or the world in black and white

In order to analyse these questions, an even simpler model was proposed, based on a binary score, taking only the values 'good' or 'bad', depending on what the co-player did when last observed[39-41]. This can be viewed as a basic system of moral assessment. In its simplest form (first order), the assessment depends only on the action of the observed player, which means on whether that player gave help or not. A discriminating donor using this assessment rule refuses help to a 'bad' recipient, and therefore becomes 'bad', which reduces the chance of being helped in turn. Effectively, discriminating players pay a cost for punishing bad co-players. Such a form of altruistic punishment is good for the community, but costly for the punisher, and thus can be viewed as a social dilemma[42]. The fact that costly punishment plays an undisputed role in other contexts, such as public goods games, ultimatum games and trust games[43-46], shows that this form of discrimination is highly plausible.

A more sophisticated assessment rule should distinguish between justified and unjustified defection, and thus take into account the score of the receiver: someone withholding help from a 'bad' player should not pay with a reduced score[35,38,39,47]. There are many assessment rules of second order, which also depend on the score of the receiver, and of third order, which depend additionally on the score of the donor. Only eight of them lead to cooperation and, at the same time, are evolutionarily stable: a homogeneous population of players using such a strategy cannot be invaded by players using other strategies[41]. All eight of these strategies distinguish justified from unjustified defections. This property must hold for any strategy which maintains cooperation, eliminates cheaters and can overcome errors[48]. The problem with the concept of justified defection, however, is that it requires information, not only about the past of the co-player, but also the past of the co-player's co-players, and their co-players etc[35,39].

Again the stability of cooperation is threatened by unconditional altruists who merely wish to keep their own good score. In a population consisting entirely of discriminators, undiscriminating co-operators fare just as well and thus spread by neutral drift. If their frequency exceeds a certain threshold, defectors can invade and take over. The situation is altered in a significant way if players occasionally fail to cooperate although their strategy prescribes it[39,49-52]. This can be due to errors in implementation or lack of resources. It is plausible that when recipients need help, donors are also short of means[52]. In this scenario, a population consisting only of conditional and unconditional altruists is not subject to random drift, but selection leads to a well defined, stable mix of the two types. This mixture can be vulnerable to the invasion of defectors[39,40].

There are several ways out of this impasse. Various assumptions on the distribution of the number of rounds lead to a bi-stable system, where depending on the initial state, the population converges either to the fixation of defectors, or to a stable mix of altruists which cannot be invaded by defectors[52,53]. In this case, the very fact that individuals are not perfect and sometimes defect involuntarily promotes the stability of cooperation[49].
Punishing a player with a bad score creates another player with a bad score; but this 'passing the buck along' can stop in two ways, by encountering either a discriminator who is uninformed, or an undiscriminating altruist. Therefore, both the lack of information and the

prevalence of unconditional altruists may, surprisingly, stabilise cooperation. A stable mix of discriminating and undiscriminating altruists which cannot be invaded by defectors, is also obtained by assuming that as players grow older, their social network grows, and with it their information about their co-players' past[54,55]. Alternatively, if the discriminating strategy distinguishes between justified and unjustified defection, the population can converge to discriminators only, which cannot be invaded by unconditional strategies[39].

Can discrimination based on the concept of 'justified defection' be destabilised by errors in perception? Not if players have the same reputation in the eyes of all members of their population. If such a consensual assessment can be achieved, the corresponding strategy is robust[38,39,41]. But if players have different views about the reputation of others, then errors in perception can undermine cooperation[56]. Such private lists of the scores of co-players are very plausible if individual interactions are only observed by a fraction of the population[35]. Gossip might be a way of achieving consensus, but can also be used for spreading false rumours and manipulating co-players. The co-evolution of human language and cooperation by indirect reciprocity is a fascinating and yet unexplored topic.

Another debated issue concerns the underlying population structure[38]. Analytical models are often based on the idealisation of a very large, well mixed population. Individual-based simulations typically assume population sizes of 50 to 100 individuals, based on estimates for the group size of hunter-gatherers. In such small populations, random drift can strongly affect evolution. Populations consisting of many separate groups with a modicum of exchange between them have also been modelled[38,53]. Such a population structure, however, facilitates the evolution of altruism via group selection[56]. It could be that co-operators fare less well than defectors within each group, but that groups of co-operators fare better than groups of defectors[57]. In extreme cases, this leads to cooperation even in the absence of indirect reciprocity. While it is most interesting to study the interaction between group selection and indirect reciprocity, it is equally important not to confuse the two effects.

Costly experiments

The basic experimental setup for testing indirect reciprocity studies a group of players equipped with an initial monetary endowment. Each player is repeatedly given the opportunity of donating money to a specific co-player, thus increasing the account of this person by an amount *b*. Players know that if they choose to do so, an amount *c* will be deducted from their own account – the cost for providing the gift. In order to eliminate confounding effects, players usually do not interact face-to-face, but they are given some information about the past actions of their potential recipients. They know that their recipients will never be their donors on future occasions, and thus that there is no scope for direct reciprocity. The interactions both with the co-players and with the experimenters are kept as anonymous as possible, usually under double-blind conditions. Many parameters can be varied within this basic setup, for instance the cost-to-benefit ratio, the size of the starting account, the number of interactions, the size of the group, the degree of information about the co-players' behaviour, the length of the game or the social background of the players.

From the first experiments onwards, it was clear that a substantial part of the players frequently decide to donate. The propensity for indirect reciprocation is apparently widespread. As expected, donations occur more frequently if the cost-to-benefit ratio is lower, or the starting account higher. Reputation has a considerable influence on the decisions. In particular, the image score of potential recipients correlates well with their expectation to

actually receive money[58]. Players who donate less often display a higher degree of discrimination. Players who are more open-handed care less about the recipient's score. Conversely, if players know that their own score is passed on, they are much more likely to donate than otherwise[59,60].

The hypothesis that more information leads to more cooperation has been confirmed in experiments which compare three information conditions[61]. In one condition, players have no information about their co-players; in the second condition, they are told about what their co-players have decided when last in the role of a donor; and in the third condition, additionally, they know about the score of the recipient of the co-player. We note that this is not always enough to decide whether a previous defection was justified or not. But the additional knowledge did enhance cooperation[61].

In this series of experiments, there is a significant positive correlation between the number of gifts given and received, but a slightly negative correlation between the number of gifts given and the total payoff obtained. In another experiment[62], however, those who give often end up with the highest payoff, so that there is a strategic advantage to generosity. The discrepancy could be due to the larger number of rounds - the advantage showed up only after a dozen of rounds - or to the fact that players were informed not only about the recipients' last move but their whole history of giving.

Evidence for strategic reputation building is found in many experiments. Donations are more frequent in earlier rounds, when the own reputation has a higher impact on the future income. But several experiments show that even players who know that their score will not be communicated show generous behaviour[61,63]. Such donors cannot be motivated by selfish interests. For many players, however, the propensity to donate more than doubles, if they know that their action will be communicated, and thus affect their own score. The influence of the recipient's score decreases accordingly. Often there is evidence for a dual motivation – players give donations if the recipient's score is high or their own score is low[35].

Experiments investigating whether a player's justified defections lowers his or her chance to receive subsequent donations indicate that cognitive problems challenge the donor[64]. Players faced with a full history of all previous rounds take in general a longer time to reach their decision, suggesting that they attempt to take into account, not only their recipient's last move, but also that of their recipient's recipient. Nonetheless, the statistics of such games with full information look surprisingly similar to those obtained when players know only the score of the recipient. Moreover, players who justifiably refuse to donate to a defector display in the following round an increased tendency to provide donations, as if to make up for that refusal. This indicates that they expect their refusal to lower their score in the co-players' eyes, and that they do not rely on the community's understanding.

Many experiments have shown that players who have just received a donation are more likely to give a donation in turn. There is evidence for 'upstream' indirect reciprocity in cyclical networks: as expected, short loops and a high benefit-to-cost ratio favour cooperation[65]. A variant of the Trust Game has two donor-responder pairs, but such that the transfers are criss-cross: the responder of one donor can only return money to the other donor (and does not know the amount transferred by that donor). The return rates turn out to be no lower than if they were addressed to the own donor[66,67]. In experimental situations which are not based on rigid networks, the decisions of donors also tend to mirror their own recent experience[58,63]. People are nicer to others if third parties were nicer to them. More generally, it seems that

decisions often depend on both the donor's payoff and the recipient's score, but such strategies have not been analysed so far.

Public goods

Indirect reciprocity is situated somewhere between direct reciprocity and public goods. On one hand, it is a game between two players only, the donor and the recipient, but on the other hand it has to be played within a larger group.

Richard Alexander claimed that indirect reciprocity originates from direct reciprocity in the presence of interested audiences[2]. A good strategy for the latter is Observer Tit For Tat[35,68]. Players using this strategy for the repeated Prisoner's Dilemma are following Tit For Tat, except that in the first round, they defect if they know that their co-player, in a previous repeated game against another player, has defected. Observer Tit For Tat relies on reputation in the first round and on personal experience in all further rounds against the same co-player. Conversely, experiments show that if several rounds of a Prisoner's Dilemma are appended to an indirect reciprocity game, the display of the previous score increases a player's probability to cooperate with generous players in the first few rounds. After a couple of rounds the personal experience obtained with the given co-player becomes more decisive[62].

The widespread tendency to judge actions between third parties, and the readiness for cooperation combined with altruistic punishment (also known as strong reciprocity[69]) has been neatly captured in an experiment involving three players[70,71]. First, players A and B engage in one round of a Prisoner's Dilemma game. Then, player C has the possibility to mete out costly punishment on A and B. Defectors are often punished although this reduces the endowment of the punisher. It would be interesting to see whether observers of a repeated Prisoner's Dilemma game are using a similar level of punishment, or whether they reduce it because wronged players have the opportunity of avenging themselves.

Indirect reciprocity and public goods games are also closely connected. For example, donors are more generous if they learn that the recipient has recently made a donation to a charitable institution[72]. An even more remarkable effect was found in an experiment alternating rounds of the public goods game with rounds of the indirect reciprocity game[73]. It is known that many players show an initial willingness to contribute to the public goods a substantial amount of their endowment, but this willingness often vanishes within a few rounds. This is not the case if indirect reciprocity games are sandwiched between the rounds of the public goods game. If players are informed about their recipient's action in the public goods game, they tend to be more generous towards recipients who contributed much. Conversely, players are more willing to contribute to the public goods if they know that this will be communicated before the start of the indirect reciprocity game. The contributions to the public good do not deteriorate from one round to the next. The donations in the indirect reciprocity game, which are channelled towards those who contributed much to the public goods, can be viewed as rewards.

Whereas most experiments, in indirect reciprocity, were motivated by models, this last experiment led to a model[74,75]. A discriminating strategy, which defects in all rounds of the indirect reciprocity game if the recipient is known to have defected in the public goods game, can guarantee stable cooperation. Since this discriminating strategy distinguishes between justified and unjustified defection, it is effectively a non-costly form of punishing free riders in the public goods game.

Cerebral expansion

Indirect reciprocity based on reputation serves as a link between diverse forms of cooperative interaction. The moralistic assessment of the other members in the population, even if they are only observed at a distance, provides a powerful tool for channelling support toward those who collaborate, and thus an incentive to join group efforts.

In human evolution, there was a tendency from the simple cooperation promoted by kin or group selection to the strategic subtleties of direct and indirect reciprocity. Direct reciprocity requires precise recognition of individual people, a memory of the various interactions one had with them in the past, and enough brain to conduct multiple repeated games simultaneously. Indirect reciprocity, in addition, has to monitor interactions among other people, possibly judge the intentions that occur in such interactions, and keep up with the ever changing social network of the group. The reputation of players may not only be determined by their own actions, but also by their associations with others.

Finally, indirect reciprocity has coevolved with human language. On one hand, it is helpful to have names for other people and to receive information about how a person is perceived by others, on the other hand a complex language is needed especially if there are intricate social interactions. The possibilities for games of manipulation, cooperation and defection are limitless. It is very likely that games of indirect reciprocity provided the selective scenario that led to cerebral expansion in human evolution.

Acknowledgments: The Program for Evolutionary Dynamics at Harvard University is sponsored by Jeffrey Epstein.

Box 1: The good, the bad and the discriminating

Similarities and differences between direct and indirect reciprocity become apparent when studying the replicator dynamics of three strategies: always cooperate, always defect and the simplest discriminating strategy. In the case of direct reciprocity, this is Tit For Tat , which helps in the first round and then does whatever the opponent did in the previous round. In the case of indirect reciprocity, this is the strategy, which prescribes to help, except if the recipient is known to have refused to help in the previous round. In each case, there are many other discriminating strategies, which are likely to take over eventually. This analysis is just a first step.

In the absence of discriminators, defectors win against cooperators. In the absence of cooperators, defectors and discriminators form a bi-stable system: depending on the initial condition, either one or the other strategy wins. In the absence of defectors, discriminators and cooperators are in equilibrium. Random fluctuations, however, make their frequencies drift up and down. The equilibria can be invaded by defectors if the frequency of discriminators is below a certain threshold. With all three strategies present, the dynamics lead either to defectors only or to a mixture of the two kinds of altruists.

If errors occur, or if an intended donation cannot be implemented through lack of resources, discriminating and undiscriminating altruists reach, in the absence of defectors, a stable co-

existence with a well-defined frequency of discriminators. If all three types are present in the population, the system displays two types of behaviour. If the frequency is too low, defectors win. If the discriminators are sufficiently frequent, all three strategies coexist. However, in the case of direct reciprocity, the frequencies oscillate periodically, whereas in the case of indirect reciprocity, they converge to an equilibrium. In each case, a long series of random fluctuations may eventually destroy the coexistence of the three strategies. In the deterministic model, only the emergence of other conditional strategies can save cooperation in the long run[76]. For stochastic population dynamics, the time average of the evolutionary oscillations can be centred on discriminators.

[Figure Box 1]

Box 2: Let a thousand morals bloom

In a world of binary moral judgements, there are four ways of assessing donors in terms of 'first order assessment': always consider them as good, always consider them as bad, consider them as good if they refuse to give, or consider them as good if they give. Only this last option makes sense. Second order assessment also depends on the score of the receiver; for example, it can be deemed good to refuse help to a bad person. There are 16 second order rules. Third order assessment also depends on the score of the donor; a good person refusing to help a bad person may remain good, for example, but a bad person refusing to help a bad person remains bad. There are 256 third order assessment rules. We display three of them. With SCORING, cooperation, C, always leads to a good reputation, G, while defection, D, always leads to a bad reputation, B. STANDING is like SCORING, but it is not bad if a good donor defects against a bad recipient. With JUDGING, in addition, it is bad to cooperate with a bad recipient.

[Figure Box 2a]

For another assessment rule, SHUNNING, all donors who meet a bad recipient become bad, regardless of what action they choose. SHUNNING strikes us as grossly unfair. Yet it emerges as the winner in a computer tournament, if errors in perception are included and if there are only a few rounds in the game[56].

An action rule for indirect reciprocity prescribes to give or to not to give, depending on the scores of both donor and recipient. For example, you may decide to help if the recipient's score is good or your own score is bad. Such an action might rise your own score and therefore increase the chance of receiving help in the future. There are 16 action rules.

If we view a strategy as the combination of an action rule and an assessment rule, we obtain 4096 strategies. In a remarkable calculation, Ohtsuki & Iwasa[41,48] analysed all 4096 strategies and proved that only 8 of them are evolutionarily stable under certain conditions, the so called 'leading eight'.

[Figure Box 2b]

The three *s in the assessment module indicate a free choice between G and B. Hence, there are 2^3 =8 different assessment rules. The action module is built as follows: if the column in the assessment module is G and B then the corresponding action is C, otherwise the action is D.

Both STANDING and JUDGING belong to the leading eight, but neither SCORING nor SHUNNING. We expect, however, that SCORING plays a similar role in indirect reciprocity as Tit-for-tat in direct reciprocity. Both strategies are not evolutionary stable, but their ability to catalyze cooperation in adverse scenarios and their simplicity constitute their strength. In extended versions of indirect reciprocity, where donors can sometimes deceive others about the reputation of the recipient, SCORING is the 'fool-proof' concept of 'I believe what I see'. SCORING judges the action and ignores the stories.

Box 3: Bidding for trust

Trust, 'a lubricant of social life'[77], plays an essential part in many types of economic transactions. In the Trust Game, donors who trust their responder will expect to gain from transferring money. In contrast, donors in the Indirect Reciprocity game know that they can expect no direct return, even if their recipient is trustworthy. All they can gain from the transfer is an increased reputation for altruism and trustworthiness.

Game theory shows that cooperation can be sustained in the indirect reciprocity game if each player carries a label[27]. The strategy prescribes that players who deviate from it have to be punished (by not being helped) for a number T of rounds. A player's label specifies for how many rounds that player has to be punished. If a donor and a receiver meet, the action prescribed by the strategy depends only on their labels, and their labels will be updated depending on the donor's action. No player has an incentive to deviate if all other players adopt this strategy, and the effect of an error will be overcome after T rounds. But settling on this strategy, for instance, on the specific number T, seems to require an institution able to guarantee honest labelling.

Subscribers to e-Bay auctions are asked to state, after every transaction, whether they were satisfied with their partner or not. Their partners' score can accordingly increase or decrease by one point. The ratings of e-Bay members, accumulated over twelve months, are public knowledge. This very crude form of assessment seems to suffice for the purpose of reputation-building, and seems to be reasonably proof against manipulation. Social history knows many other instances of public score-keeping: 'Societies have sewn scarlet letters to people's garments, shaved heads, cut off fingers and given medals to signal to strangers some aspect of an individual's past deeds or misdeeds.'[29] Reputation mechanisms have also played an important role in the emergence of medieval trade[78].

Box 4: Games of direct reciprocity

Trust Game: two players, one in the role of the donor, the other in the role of the responder. The donor can transfer some money to the responder. Upon arrival, the amount is multiplied by three. The responder, then, has the possibility of sending some of it back to the donor. A responder with an income-maximising strategy should send nothing back. Any donor expecting this should therefore transfer nothing. In real experiments, many donors transfer substantial amounts, and some obtain large returns, so that both players win.

Public Goods: each of N players can independently decide to transfer some money to a common pool, where it is multiplied by some factor r (smaller than N) and then divided equally among all players irrespective of whether they have contributed or not. Since each player receives, in return for the own contribution, only the fraction r/N, the income-maximising strategy is to contribute nothing. In real experiments, however, many players contribute. If all do, they multiply their endowment by the factor r.

The Public Goods game for N=2 players has the structure of a Prisoner's Dilemma. Two players who cooperate earn more than two players who defect; but a defector cheating on a cooperator earns the highest payoff, and the exploited cooperator earns the lowest. If two players, in a trust game, are simultaneously in the role of the donor, and then simultaneously in the role of the responder, they play two rounds of a Prisoner's Dilemma.

Experimental economists, and experimental psychologists, have intensively studied these games, and diverse variations[79,80].

Box 5: Upstream and downstream

An entirely different approach to indirect reciprocity is based, not on the reputation of the recipient, but on the experience of the donor. An individual who has been at the receiving end of a donation may feel motivated to donate in turn. This has been called 'upstream' reciprocity, in contrast to the 'downstream' reciprocity that we have considered here[2]. In both cases, the decision to give help can be interpreted as a misdirected act of gratitude. In one case, recipients are thanked for what another did. In the other case, they are thanked by someone who did not profit by what they did. Upstream reciprocity seems more liable to exploitation, since the support is not directed towards those who gave support. If players are arranged in closed loops, for instance: A gives B, who gives C, who gives D, who gives A again, cooperation can be stable, provided the loops are very short and the number of rounds sufficiently high[81,82]. If the players meet randomly, however, or the social network is fluid, then upstream reciprocity is too diffuse to prevent defectors from taking over[2,55].

[Figure Box 5].

Box 6: Social viscosity

Altruism toward genetic relatives can evolve by kin selection provided Hamilton's rule[14] holds: the coefficient of relatedness, r, between the donor and the recipient has to exceed the cost to benefit ratio of the altruistic act,

$$r>c/b.$$

As Haldane has said, 'I will jump into the river to save two brothers or eight cousins.' The probability that brothers share a 'selfish gene' is ½; the same probability for cousins is 1/8. Kin selection works in 'viscous' populations where chances are high that neighbours are genetic relatives.

For indirect reciprocity a similar rule holds[35]: the probability, q, to know the social score of another person must exceed the cost to benefit ratio

$$q>c/b.$$

The role of genetic relatedness that is crucial for kin selection is replaced by social acquaintanceship. In a fluid population, where most interactions are anonymous and people have no possibility of monitoring the social score of others, indirect reciprocity has no chance. In a socially viscous population, where people know each others reputation, cooperation by indirect reciprocity can thrive.

References

1. Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35-57 (1971).

2. Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).

3. Binmore, K. G. *Game Theory and the Social Contract* (MIT Press, Cambridge, Massachusetts, 1994).

4. Seabright, P. *The Company of Strangers: A Natural History of Economic Life* (Princeton Univ. Press, Princeton, 2004).

5. Bolton, G. E., Katok, E. & Ockenfels, A. How effective are electronic reputation mechanisms? An experimental investigation. *Manage. Sci.* **50**, 1587-1602 (2004).

6. Keser, C. Trust and reputation building in e-commerce. IBM Watson Research Center, CIRANO working paper No. 2002s-75 <http://www.cirano.qc.ca/en/publication_detail.php?id=2002s-75> (2002).

7. Dellarocas, C. Efficiency and robustness of binary feedback mechanisms in trading environments with moral hazard. MIT Sloan School of Management working paper No. 4297-03 <http://ssrn.com/abstract=393043> (2003).

8. Bolton, G., Katok, E. & Ockenfels, A. Bridging the trust gap in electronic markets: A strategic framework for empirical study. in *Applications of Supply Chain Management and e-Commerce Research* (eds Geunes, J., Akcali, E., Pardalos, P. M., Romeijn, H. E. & Shen, Z. J.) 195-216 (Springer, formerly Kluwer Academic Publishers, Dordrecht, 2004).

9. Resnick, P., Zeckhauser, R., Friedman, E. & Kuwabara, K. Reputation systems. *Commun. ACM* **43**, 45-48 (2000).

10. Lucking-Reiley, D., Bryan, D., Prasad, N. & Reeves, D. Pennies from eBay: The determinants of price in online auctions. Vanderbilt Univ., Univ. Arizona working paper (1999).

11. Maynard Smith, J. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ. Press, Oxford, 1997).

12. Wilson, E. O. *Sociobiology* (Harvard Univ. Press, Cambridge, Massachusetts, 1975).

13. Trivers, R. *Social Evolution* (Benjamin Cummings, Menlo Park, 1985).

14. Hamilton, W. D. *Narrow Roads of Gene Land* Vol. 1 (Freeman, New York, 1996).

15. Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E. & Cohen, J. D. The neural basis of economic decision-making in the ultimatum game. *Science* **300**, 1755-1758 (2003).

16. Rilling, J. K. *et al.* A neural basis for social cooperation. *Neuron* **35**, 395-405 (2002).

17. Fehr, E. The neural basis of altruistic punishment. *Science* **305**, 1254-1258 (2004).

18. Dunbar, R. *Grooming, Gossip and the Evolution of Language* (Harvard Univ. Press, Cambridge, Massachusetts, 1996).

19. Brown, D. E. *Human Universals* (McGraw-Hill, New York, 1991).

20. Whiten, A. & Byrne, R. W. (eds) *Machiavellian Intelligence II: Extensions and Evaluations* (Cambridge Univ. Press, Cambridge, UK, 1997).

21. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984). (reprinted in Penguin, Harmondsworth, 1989).

22. Nowak, M. A. & Sigmund, K. The alternating prisoner's dilemma. *J. Theor. Biol.* **168**, 219-226 (1994).

23. Frean, M. R. The prisoner's dilemma without synchrony. *P. Roy. Soc. Lond. B Bio.* **257**, 75-79 (1994).

24. Berg, J., Dickhaut, J. & McCabe, K. Trust, reciprocity and social history. *Game. Econ. Behav.* **10**, 122-142 (1995).

25. Fudenberg, D. & Maskin, E. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* **50**, 533-554 (1986).

26. Rosenthal, R. W. Sequences of games with varying opponents. *Econometrica* **47**, 1353-1366 (1979).

27. Kandori, M. Social norms and community enforcement. *Rev. Econ. Stud.* **59**, 63-80 (1992).

28. Ellison, G. Cooperation in the prisoner's dilemma with anonymous random matching. *Rev. Econ. Stud.* **61**, 567-588 (1994).

29. Okuno-Fujiwara, M. & Postlewaite, A. Social norms in matching games. *Game. Econ. Behav.* **9**, 79-109 (1995).

30. Nowak, M. A. & Sigmund, K. Evolutionary dynamics of biological games. *Science* **303**, 793-799 (2004).

31. Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250-253 (1992).

32. Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646-650 (2004).

33. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390-1396 (1981).

34. Nowak, M. A. & Sigmund, K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner’s dilemma game. *Nature* **364**, 56-58 (1993).

35. Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573-577 (1998).

36. Wedekind, C. Give and ye shall be recognised. *Science* **280**, 2070-2071 (1998).

37. Ferrière, R. Help and you shall be helped. *Nature* **393**, 517-519 (1998).

38. Leimar, O. & Hammerstein, P. Evolution of cooperation through indirect reciprocation. *P. Roy. Soc. Lond. B Bio.* **268**, 745-753 (2001).

39. Panchanathan, K. & Boyd, R. A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *J. Theor. Biol.* **224**, 115-126 (2003).

40. Nowak, M. A. & Sigmund, K. The dynamics of indirect reciprocity. *J. Theor. Biol.* **194**, 561-574 (1998).

41. Ohtsuki, H. & Iwasa, Y. How should we define goodness? Reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **231**, 107-120 (2004).

42. Dawes, R. M. Social dilemmas. *Annu. Rev. Psychol.* **31**, 169-193 (1980).

43. Fehr, E. & Gächter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980-994 (2000).

44. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137-140 (2002).

45. Fehr, E. & Fischbacher, U. The nature of human altruism. *Nature* **425**, 785-791 (2003).

46. Sigmund, K., Hauert, C., & Nowak, M. A. Reward and punishment. *P. Natl. Acad. Sci. USA* **98**, 10757-10762 (2001).

47. Sugden, R. *The Economics of Rights, Cooperation and Welfare* (Basil Blackwell, Oxford, 1986).

48. Ohtsuki, H. & Iwasa, Y. The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. theor. Biol. In press* (2005)

49. Lotem, A., Fishman, M. A. & Stone, L. Evolution of cooperation between individuals. *Nature* **400**, 226-227 (1999).

50. Lotem, A., Fishman, M. A. & Stone, L. Evolution of unconditional altruism through signalling benefits. *P. Roy. Soc. Lond. B Bio.* **270**, 199-205 (2002).

51. Fishman, M. A., Lotem, A. & Stone, L. Heterogeneity stabilises reciprocal altruism interaction. *J. Theor. Biol.* **209**, 87-95 (2001).

52. Fishman, M. A. Indirect reciprocity among imperfect individuals. *J. Theor. Biol.* **225**, 285-292 (2003).

53. Brandt, H. & Sigmund, K. The logic of reprobation: Assessment and action rules for indirect reciprocity. *J. Theor. Biol.* **231**, 475-486 (2004).

54. Mohtashemi, M. & Mui, L. Evolution of indirect reciprocity by social information: The role of trust and reputation in evolution of altruism. *J. Theor. Biol.* **223**, 523-531 (2003).

55. Brandt, H. & Sigmund, K. Indirect reciprocity, image scoring, and moral hazard. *P. Natl. Acad. Sci. USA* **102**, 2666-2670 (2005).

56. Takahashi, N. & Mashima, R. The emergence of indirect reciprocity: Is the standing strategy the answer? Hokkaido Univ. working paper No. 29 <http://lynx.let.hokudai.ac.jp/COE21/pdf/029.zip> (2003).

57. Wilson, D.S. A theory of group selection. *Proc. Natl. Acad. Sc USA* **72**,143-146 (1975).

58. Wedekind, C. & Milinski, M. Cooperation through image scoring in humans. *Science* **288**, 850-852 (2000).

59. Seinen, I. & Schram, A. Social status and group norms: Indirect reciprocity in a helping experiment. CREED, Univ. Amsterdam working paper (2001).

60. Seinen, I. & Schram, A. Social status and group norms: Indirect reciprocity in a repeated helping experiment. *Eur. Econ. Rev.* in press (2004).

61. Bolton, G. E., Katok, E. & Ockenfels, A. Cooperation among strangers with limited information about reputation. *J. Public Econ.* in press (2004).

62. Wedekind, C. & Braithwaite, V. A. The long-term benefits of human generosity in indirect reciprocity. *Curr. Biol.* **12**, 1012-1015 (2002).

63. Engelmann, D. & Fischbacher, U. Indirect reciprocity and strategic reputation building in an experimental helping game. Univ. Zürich working paper No. 132 <http://www.iew.unizh.ch/wp/iewwp132.pdf> (2002).

64. Milinski, M., Semmann, D., Bakker, T. C. M. & Krambeck, H. J. Cooperation through indirect reciprocity: Image scoring or standing strategy? *P. Roy. Soc. Lond. B Bio.* **268**, 2495-2501 (2001).

65. Greiner, B. & Levati, M. V. Indirect reciprocity in cyclical networks: An experimental study. MPI Jena working paper No. 15 <ftp://papers.mpiew-jena.mpg.de/esi/discussionpapers/2003-15.pdf> (2003).

66. Dufwenberg, M., Gneezy, U., Güth, W. & van Damme, E. Direct vs indirect reciprocity: An experiment. *Homo Oeconomicus* **18**, 19-30 (2001).

67. Güth, W., Königstein, M., Marchand, N. & Nehring, K. Trust and reciprocity in the investment game with indirect reward. *Homo Oeconomicus* **18**, 241-262 (2001).

68. Pollock, G. B. & Dugatkin, L. A. Reciprocity and the evolution of reputation. *J. Theor. Biol.* **159**, 25-37 (1992).

69. Gintis, H. *Game Theory Evolving: A Problem-Centered Introduction to Modeling Stategic Interaction* (Princeton Univ. Press, Princeton, 2000).

70. Fehr, E. & Fischbacher, U. Social norms and human cooperation. *Trends Cogn. Sci.* **8**, 185-190 (2004).

71. Fehr, E. & Fischbacher, U. (2004) Third party punishment and social norms. *Evol. Hum. Behav.* **25**, 63-87 (2004).

72. Milinski, M., Semmann, D. & Krambeck, H. J. Donors in charity gain in both indirect reciprocity and political reputation. *P. Roy. Soc. Lond. B Bio.* **269**, 881-883 (2002).

73. Milinski, M., Semmann, D. & Krambeck, H. J. Reputation helps solve the 'tragedy of the commons'. *Nature* **415**, 424-426 (2002).

74. Panchanathen, K. & Boyd, R. Indirect reciprocity can stabilise cooperation without the second-order free-rider problem. *Nature* **432**, 499-502 (2004).

75. Fehr, E. Don't lose your reputation. *Nature* **432**, 449-450 (2004).

76. Brandt, H. & Sigmund, K. The good, the bad and the discriminator. *J. Theor. Biol.* to appear (2005).

77. Arrow, K. *The Limits of Organization* (Norton, New York, 1974).

78. Milgrom, P., North, D. & Weingast, B. The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Economics & Politics* **2**, 1-23 (1990).

79. Camerer, C. E. *Behavioral Game Theory* (Princeton Univ. Press, Princeton, 2003).

80. Colman, A. M. *Game Theory and its Applications in the Social and Biological Sciences* (Butterworth-Heinemann, Oxford, 1995).

81. Boyd, R. & Richerson, P. J. The evolution of indirect reciprocity. *Soc. Networks* **11**, 213-236 (1989).

82. Pfeiffer, T., Rutte, C., Killingback, T., Taborsky, M., Bonhoeffer, S. Evolution of cooperation by generalized reciprocity. *Proc. R. Soc. B* in press

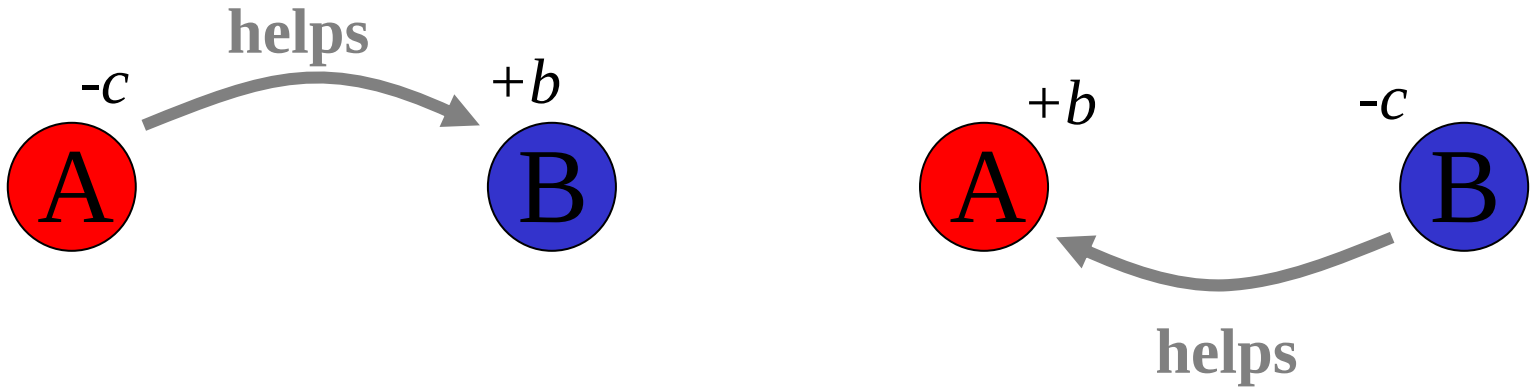
Figure legends

Figure 1: Direct reciprocity means that A helps B and later B helps A. Indirect reciprocity means that A helps B and later C helps A. Helping represents a cost, *c*, for the donor and a benefit, *b*, for the recipient. Direct reciprocity is captured in the framework of the repeated Prisoner's Dilemma. Indirect reciprocity, in contrast, is built on reputation. Helping another individual is costly but buys the reputation of being helpful. Computer simulations and mathematical investigations of indirect reciprocity show that natural selection can favour strategies that help others based on their reputation.

Figure 2: In a natural extension of the basic model of indirect reciprocity, an action between a donor, A, and a recipient, B, is observed by a subset of the population[35]. The observers, the donor and the recipient can inform others about what has happened. There are many possibilities of error: the action or the intention of the donor can interpreted differently by different people, some individuals may receive conflicting information from different sources, some individuals may not receive any information at all. Therefore, the reputation of a person is not simply a label that is visible to all others, but instead each person has a private list of the reputation of others. While language might help to compare and synchronize these lists[41], ultimately reputation is in the eyes of the beholder.

Figure 1

a. Direct reciprocity



b. Indirect reciprocity

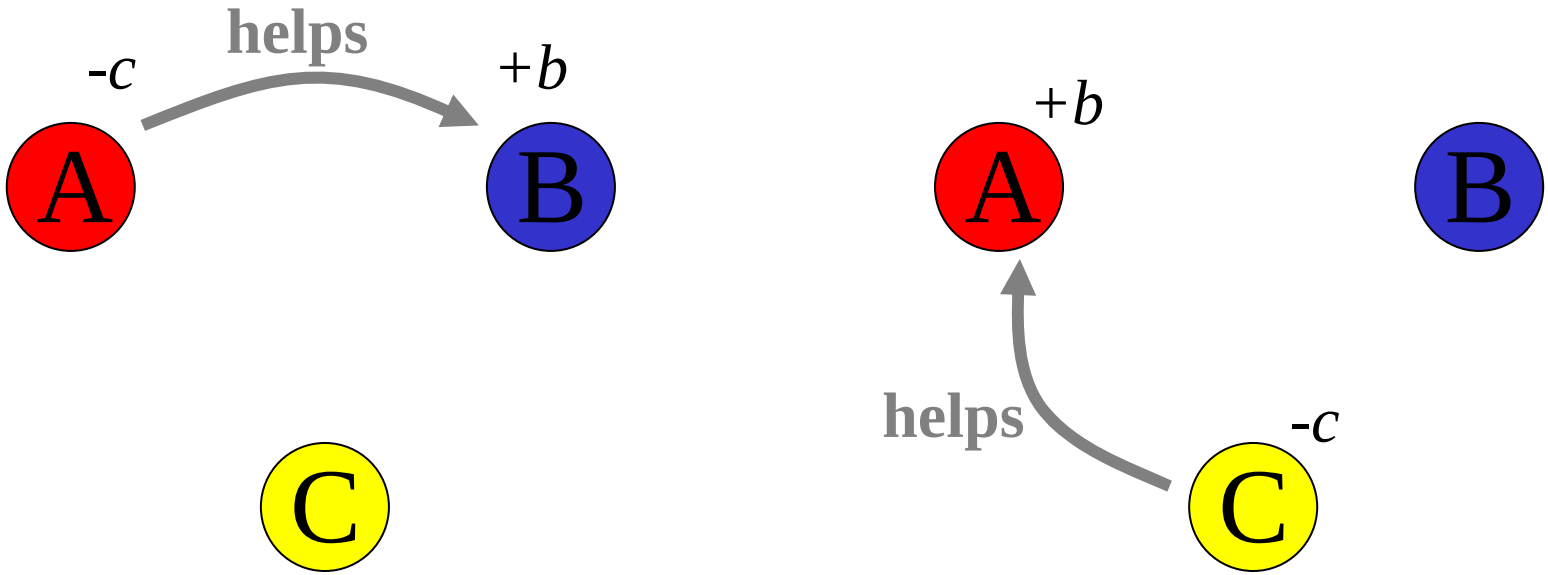


Figure 2

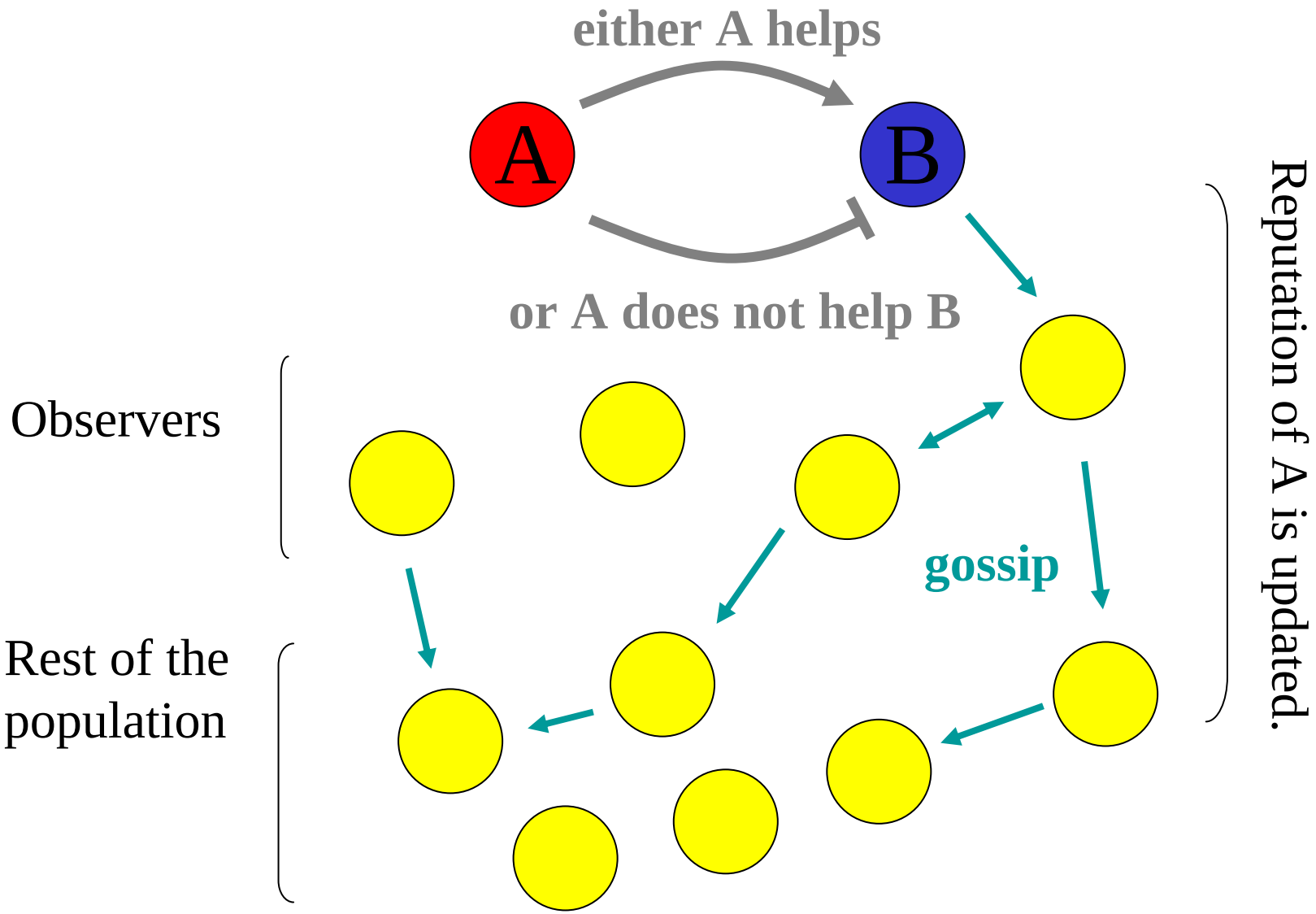

either A helps
A
B
or A does not help B
Observers
Rest of the population
gossip
Reputation of A is updated.

Figure Box 1

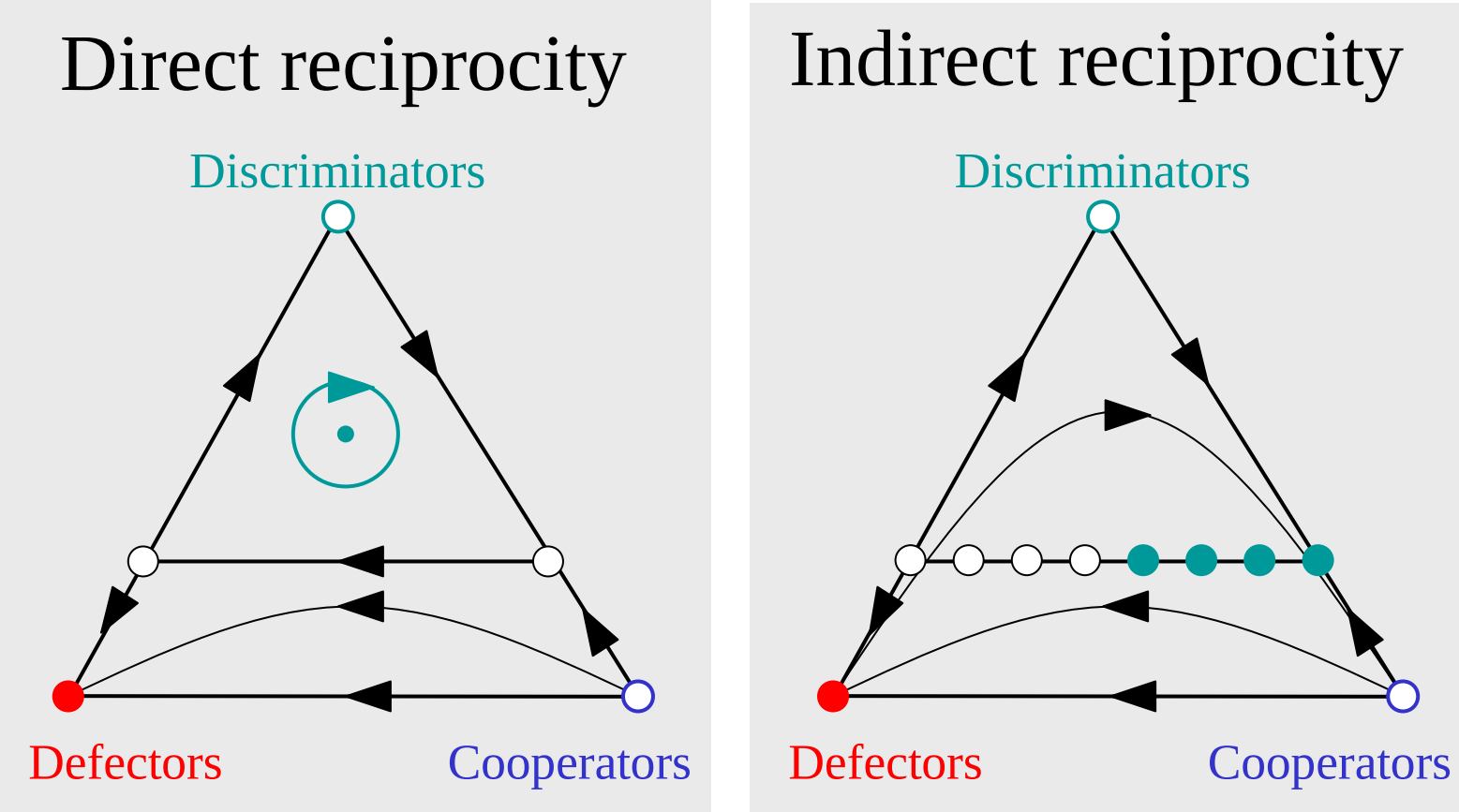

Direct reciprocity
Discriminators
Defectors
Cooperators
Indirect reciprocity
Discriminators
Defectors
Cooperators

Figure Box 2a

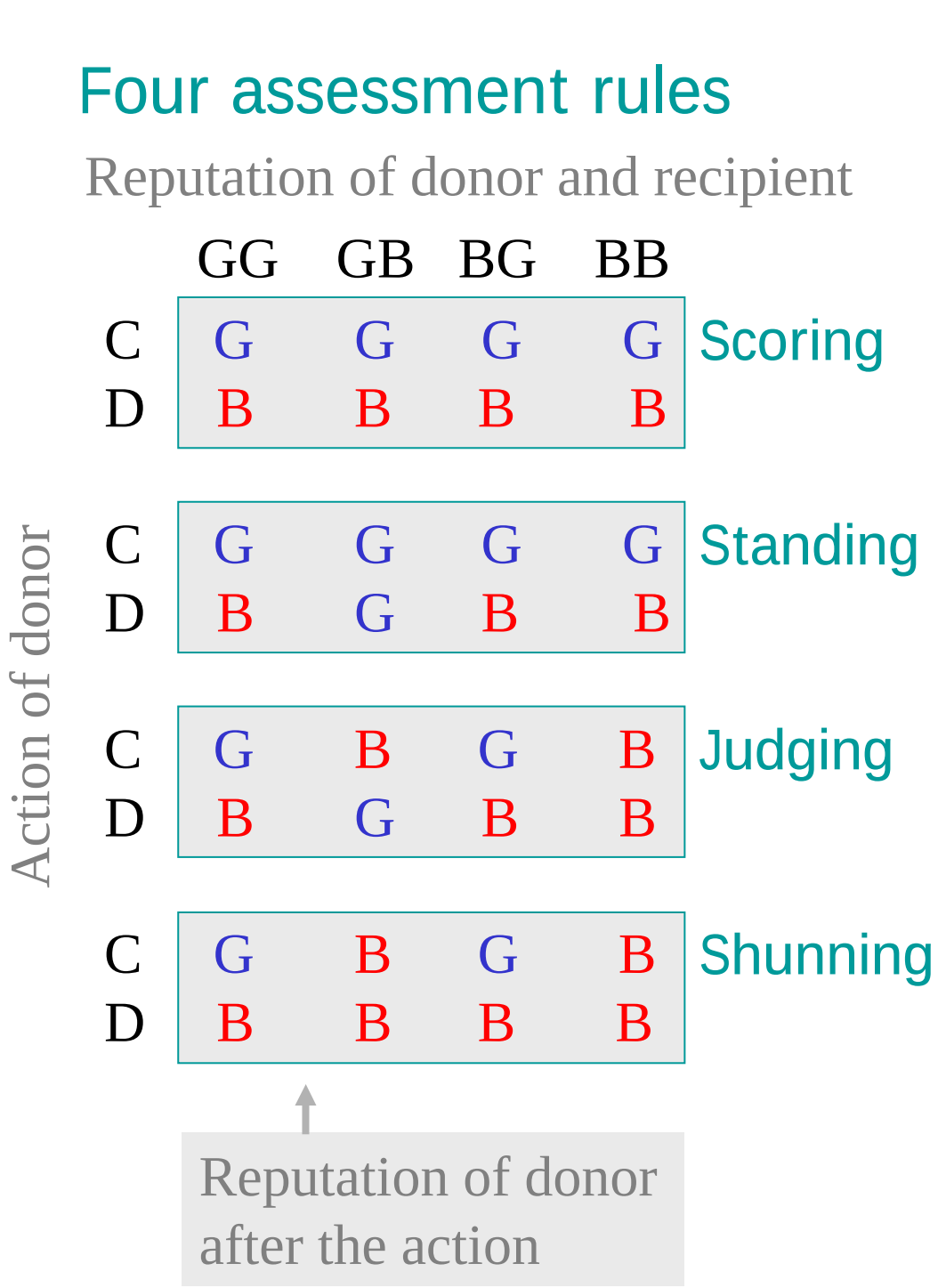
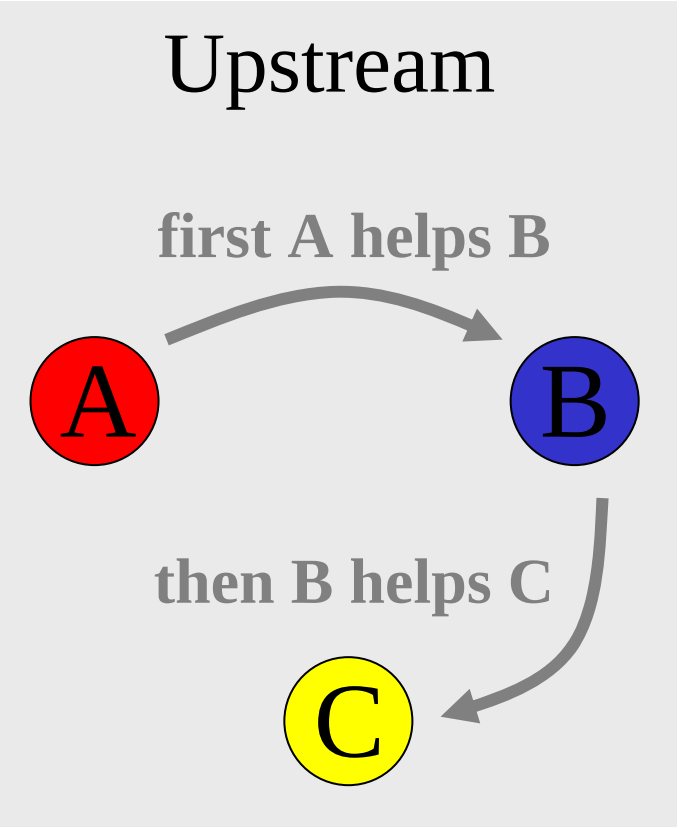

Four assessment rules
Reputation of donor and recipient
GG GB BG BB
Action of donor
C G G G G Scoring
D B B B B
C G G G G Standing
D B G B B
C G B G B Judging
D B G B B
C G B G B Shunning
D B B B B
Reputation of donor after the action

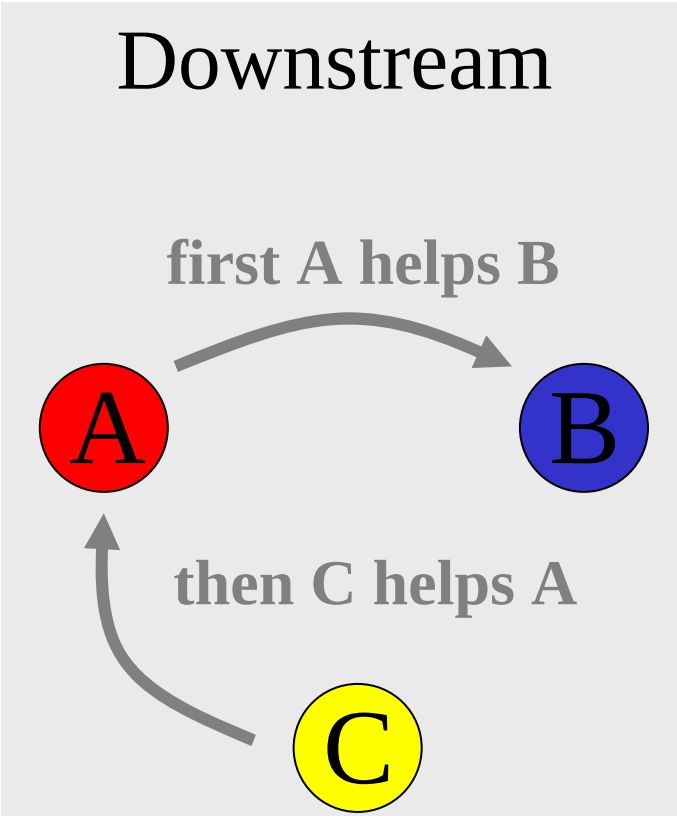
Figure Box 2b

Ohtsuki & Iwasa's *Leading eight*

	GG	GB	BG	BB	
C	G	*	G	*	Assessment
D	B	G	B	*	
	C	D	C	C/D	Action

Figure Box 5


Upstream
first A helps B
A
B
then B helps C
C


Downstream
first A helps B
A
B
then C helps A
C